



Meng, F., Yan, S., Nikolovgenis, K., Bi, Y., Ou, Y., Wang, R., Nejabati, R., & Simeonidou, D. (2018). Field Trial of Gaussian Process Learning of Function-Agnostic Channel Performance Under Uncertainty. In *2018 Optical Fiber Communications Conference and Exposition (OFC)* <https://doi.org/10.1364/OFC.2018.W4F.5>

Peer reviewed version

Link to published version (if available):  
[10.1364/OFC.2018.W4F.5](https://doi.org/10.1364/OFC.2018.W4F.5)

[Link to publication record in Explore Bristol Research](#)  
PDF-document

This is the author accepted manuscript (AAM). The final published version (version of record) is available online via OSA at <https://doi.org/10.1364/OFC.2018.W4F.5> . Please refer to any applicable terms of use of the publisher.

## University of Bristol - Explore Bristol Research

### General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:  
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>

# Field Trial of Gaussian Process Learning of Function-Agnostic Channel Performance Under Uncertainty

F. Meng, S. Yan, K. Nikolovgenis, Y. Ou, R. Wang, Y. Bi, E. Hugues-Salas, R. Nejabati, D. Simeonidou  
*High Performance Networks Group, University of Bristol, UK*  
*f.meng@bristol.ac.uk*

**Abstract:** We model and experimentally demonstrate a novel performance learning method based on monitoring and Gaussian process. After 436km dark fiber transmission the model captures most of the test data with reasonable prediction error and enables a robust QoT predictor.

**OCIS codes:** (060.1155) All-optical networks; (060.4256) Networks optimization

## 1. Introduction

Optical performance monitoring has been extensively applied in optical networks not only reactively where network restoration is triggered after failure detection[1], but also proactively where monitoring is treated as a data source to train the system model and further predict future services' quality-of-transmission (QoT)[2]. One difficulty in predicting other channels' QoT by learning from the existing channels is model selection, common sense would be regression in a weight-space view[3]. However, the Erbium-doped fiber amplifiers (EDFA) gain spectrum and noise figure are wavelength( $\lambda$ ) dependent while the link loss is wavelength independent, hence channels with distant wavelengths are less noise-representative than neighbor ones. This gain/loss wavelength discrepancy results in severe optical-signal-to-noise ratio (OSNR) non-uniformity across the C-band after passing through a cascade of EDFA's[4]. Also, each EDFA is expected to have some level of undetected noise perturbation, and each fiber span would give uncertain power loss due to ageing or poor connections[2], the compound effect on OSNR performance will vary among channels that is very hard to parameterize, especially in the case of sparse channel distributions because of add-drops. Given there is no prior knowledge of what the prediction model is, i.e. function-agnostic, any arbitrarily chosen model will result in under-fitting or over-fitting issues which further leads to network failure or margin over-provisioning due to poor QoT estimations[5]. In contrast to the weight-space view method, Gaussian Process (GP) is a stochastic probability distribution over functions (function-space view), any inference takes place directly in the space of functions that derived from the data. So rather than claiming the optical link model (noise vs  $\lambda$ ) to be linear, cubic, etc. GP can represent the model obliquely, but also rigorously by letting the monitoring data "speak" more[6].

In this paper, we propose, model, and experimentally demonstrate Gaussian process regression (GPR) for OSNR performance inference under system uncertainties. In this region-wide field trial, performance monitoring is used for model training, the result shows that GP performs better than other prediction methods with a mean error of 0.7dB.

## 2. Noisy Gaussian process inference

Fig.1(a) shows the flow chart of the overall learning algorithm, the controller updates its database and goes through the process every time a new channel is lit. To fit in a GP, the monitoring data (OSNR vs  $\lambda$ ) of the existing channels is seen as the training set, a new channel's OSNR performance is seen as the test set. Since the data is intrinsically noisy in which the monitored OSNR fluctuates around the true value over time, GP models additive independent and identically distributed (iid) Gaussian noise  $\epsilon \sim N(0, \sigma_Q^2)$  to the monitoring data such that  $OSNR(\text{monitored}) = Q(\lambda) + \epsilon$  where  $Q(\lambda)$  is the true theoretical QoT function. Hence all future inferences are made by taking the noise variance into account. Fig.1(b) shows the graphical model for GP, the inputs  $\lambda_i$  and outputs  $Q_i$  (monitored OSNR) of the training set are known data through which the function node ( $f_i$ ) is unknown. Each monitored data  $Q_i$  is conditionally independent of all other nodes given the latent variable  $f_i$ . To predict a new test channel  $Q_*$ , GP samples functions for  $f_*$  that is conditioned on  $\lambda_*$  and the training set. We model the similarity kernel (covariance function) using "squared

exponential"  $k(\lambda, \lambda') = \sigma_f^2 \exp\left[\frac{-(\lambda - \lambda')^2}{2l^2}\right] + \sigma_Q^2 I$  where  $\sigma_f^2$  and  $l$  are hyperparameters that affect the shape and smoothness of GPR[6]. The kernel function measures how similar two points are correlated to each other, for example if two wavelengths are close to each other, their OSNR performance should be nearly identical ( $k$  reaches maximum) given their lightpath is the same. Then the joint multivariate Gaussian distribution of the monitoring data  $Q$  and the test data  $Q_*$  according to the prior (initial believe of the hidden function  $f_i$ ) is  $\begin{bmatrix} Q \\ Q_* \end{bmatrix} \sim N(0, \begin{bmatrix} K(\lambda, \lambda) + \sigma_Q^2 I & K(\lambda, \lambda_*) \\ K(\lambda_*, \lambda) & K(\lambda_*, \lambda_*) \end{bmatrix})$

where  $K$  is a multi-dimensional matrix which is determined by the input data. The posterior OSNR estimation of the test set  $Q_*$  conditioned on the training set  $\lambda$ ,  $Q$  and test input  $\lambda_*$  follows a Gaussian distribution:

$$Q_* | Q, \lambda, \lambda_* \sim N(K(\lambda_*, \lambda)(K(\lambda, \lambda) + \sigma_Q^2 I)^{-1} Q, K(\lambda_*, \lambda_*) - K(\lambda_*, \lambda)(K(\lambda, \lambda) + \sigma_Q^2 I)^{-1} K(\lambda, \lambda_*))$$

Fig. 1(c) shows the pseudocode that summarizes the GP algorithm. The hyperparameters are optimised by maximising the marginal likelihood (maximum likelihood) which is by means of seeking partial derivatives w.r.t.  $\sigma_f^2$  and  $l[6]$ .

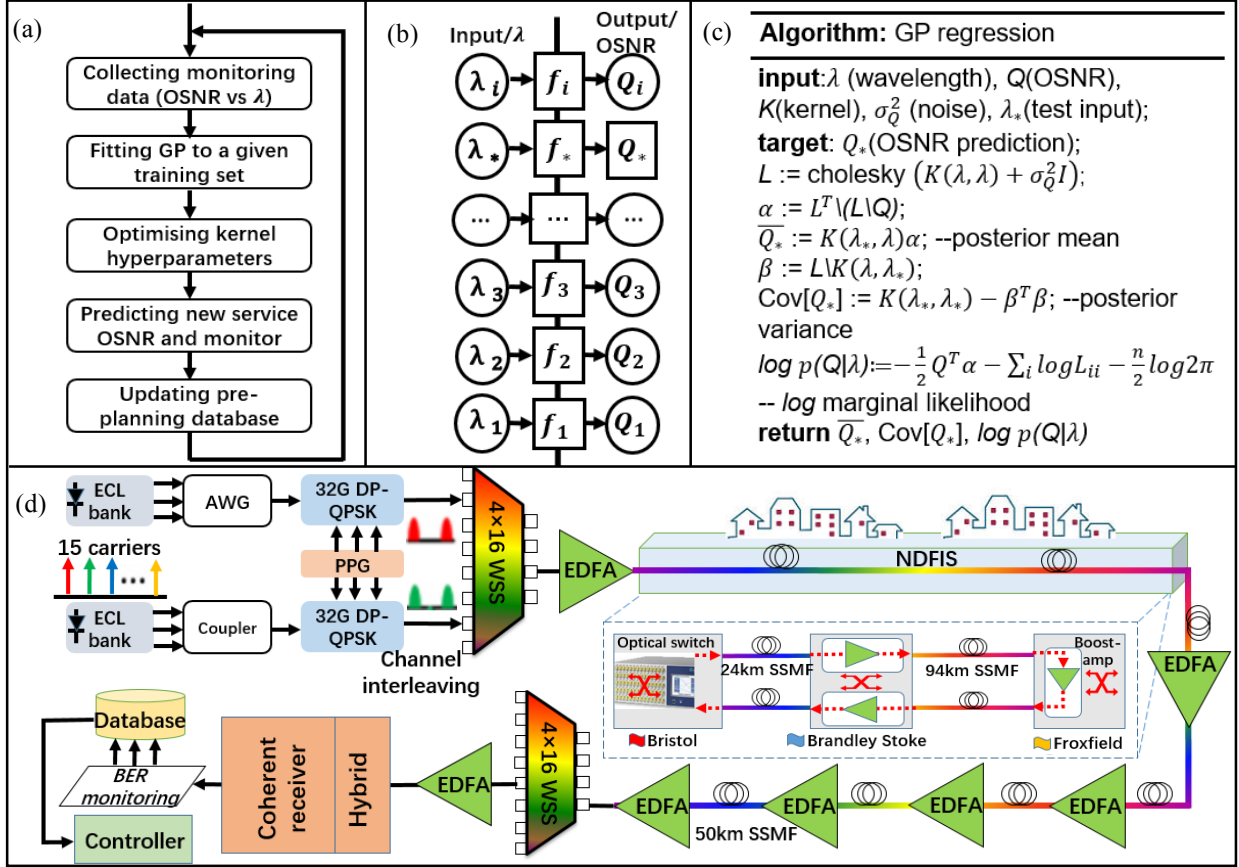


Fig.1 (a) Controller algorithm flow chart, (b) principle of GP, (c) GP algorithm pseudocode, (d) field trial UK testbed with NDFIS, ECL(external cavity laser), PPG(pulse pattern generator), DP-QPSK(dual polarization quadrature phase shift keying)

### 3. Field trial setup and result analysis

Fig. 1(d) shows the field trial testbed setup using part of the UK's National Dark Fiber Infrastructure Service (NDFIS) which connects three geographical nodes: University of Bristol, Brandley Stoke and Froxfield [http://www.ndfis.org/]. 15 equalized 50GHz-spaced 32Gbaud DP-QPSK signals (training data, wavelength ranging from 1545nm to 1565nm) are generated at the transmitter side (Fig. 2(a)), a Wavelength Selective Switch (WSS) is used both as a filter and interleaver to avoid crosstalk. All the other 32 wavelengths within the wavelength range are treated as test channels. Launch power is set to 0 dBm/channel/span, launch OSNR is kept identical among the 15 channels by conducting back-to-back Error-Vector-Magnitude(EVM) based bit-error-rate(BER) monitoring. The NDFIS loop-back link gives 236km effective transmission length, at each site a boost amplifier is used to completely compensate the span loss. Another 200km fiber link is added after the loop-back (giving 436km in total) where signals are amplified every 50km before being coherently received. Some of the pre-tested EDFAs introduce unexpected noise because of ageing. OSNRs are calculated by averaging the BERs over 5minutes/channel period using  $OSNR_{rx} = (\text{erfc}^{-1}(2BER_{mean}))^2 * 2Rs/Bn$  where  $Rs$  and  $Bn$  are the signal baudrate and noise level bandwidth respectively.

Fig. 2(b) shows the receiver side constellation diagrams of four training channels with different  $\lambda$ s, the launch (back-to-back) signal performances are almost identical, while after transmission their performances show different degradations. Fig. 2(c) is the GPR curve which represents the posterior mean estimation of the test data given all the training data. We model the monitoring noise/variance  $\sigma_Q^2 = 0.5\text{dB}$  to withstand OSNR discrepancy. The GP curve does not necessarily pass through each training point but is always within the variance range. The curve shape depends on the kernel/covariance function hyperparameters  $\sigma_f^2$  and  $l$  which in this case, are 2.07 and 1.53 respectively. The shaded area represents pointwise 95% confidence interval of the prediction which is computed by  $\bar{Q}_* \pm 1.96\sqrt{\text{Var}(Q_*)}$ . The confidence integral indicates the posterior prediction uncertainty which goes low where training data is enough, and goes high where there is no training data. Channels with  $\lambda$  around 1557nm present the best OSNR performance while  $\lambda$  of the worst OSNR performance locates around 1548nm due to unpredictable accumulated noise. From the

distribution of all the test channel OSNR performance (the monitoring data), only 4 points fall outside of the confidence integral while 88% of the points fall within it. To evaluate the performance of GP, two other estimation methods are used, (1) least-square linear regression (LSLR): the overall fitted line minimises the sum of squares of residuals/errors, (2) neighbor-average(NA): averaging OSNRs of the neighbor training channels sitting at both sides. The linear regression line is plotted in fig. 2(c) with gradient 0.3048. Fig. 2(d) shows the test set prediction errors of the three methods, the error is defined as the absolute value subtracting the monitoring data from the prediction data. GP gives the lowest maximum (MAX) error of 1.2dB (also shown in fig. 2(e)), compared to 2.2dB of NA and 4.5dB of LSLR. Fig. 2(e) further shows the root mean square deviation (RMSD) of the OSNR prediction errors, GP outputs the lowest RMSD value of 0.7dB compared to others, demonstrating an enhanced QoT predictor.

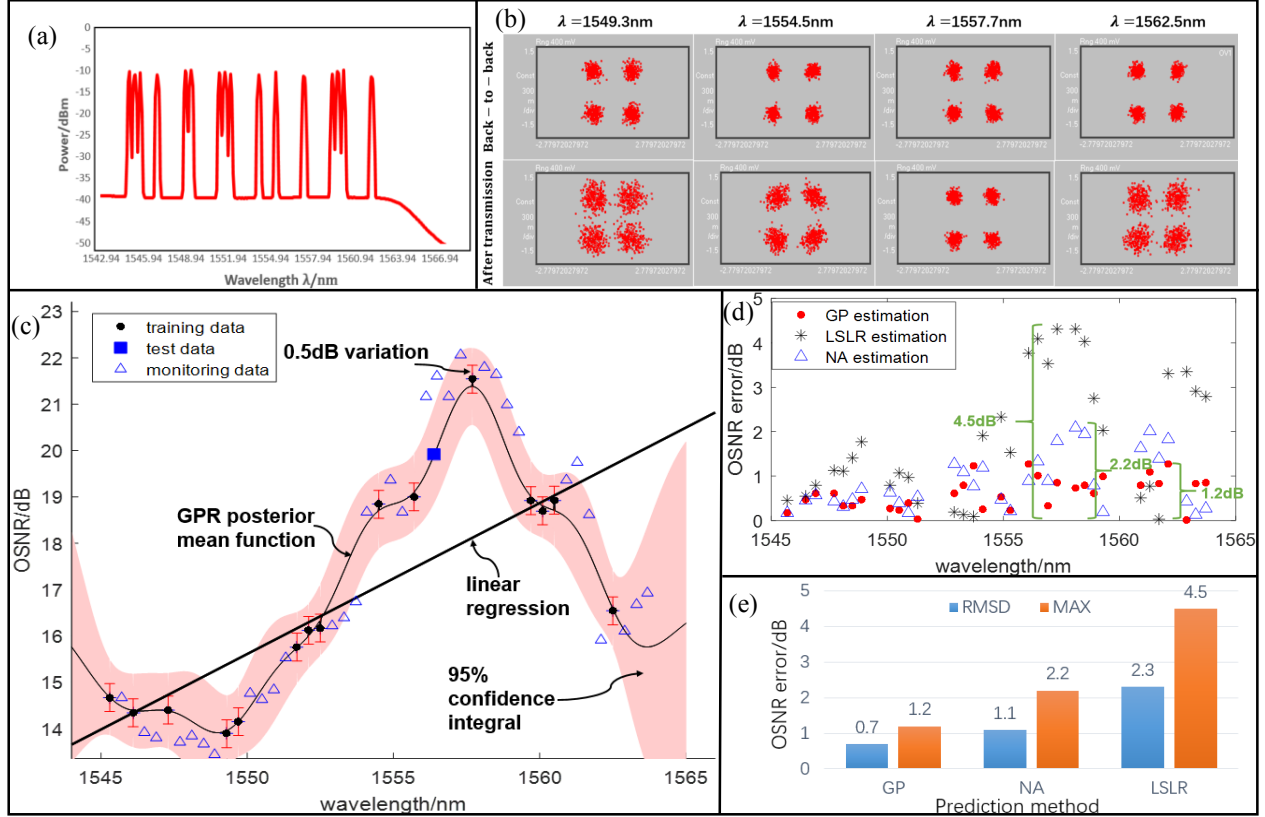


Fig.2(a) 15 equalized training spectrum and one test spectrum, (b) constellation diagrams of four  $\lambda$ s before and after transmission, (c) GP learning curve plot conditioned on the training set with optimized hyperparameters, (d) prediction error performance plot of the three methods, (e) RMSD and MAX error plot of the three methods

#### 4. Summary

We have experimentally demonstrated in this field trial a novel QoT learning method using performance monitoring and Gaussian processes which outperforms other learning methods. A mean-squared prediction error of 0.7dB is achieved with GP over a broad range of wavelength and without any prior system knowledge. This precise QoT prediction capability is ready to be integrated into controllers to enable a proactive “self-learning” network that runs close to its performance limit and saves significant margins under noise uncertainties.

#### Acknowledgement

This work is supported by EPSRC grant EP/L020009/1: TOUCAN and EP/L026155/2: INSIGHT project.

#### References

- [1] F. Meng *et al.*, "Field trial of a novel SDN enabled network restoration utilizing in-depth optical performance monitoring assisted network re-planning," *2017 Optical Fiber Communications Conference and Exhibition (OFC)*, Los Angeles, CA, 2017, pp. 1-3.
- [2] F. Meng *et al.*, "Robust Self-learning Physical Layer Abstraction Utilizing Optical Performance Monitoring and Markov Chain Monte Carlo," *2017 European Conf. and Exhibition on Optical Communication (ECOC)*, Gothenburg, Sept. 2017, paper W.3.A.2.
- [3] P. Samadi *et al.*, "Quality of Transmission Prediction with Machine Learning for Dynamic Operation of Optical WDM Networks," *2017 European Conf. and Exhibition on Optical Communication (ECOC)*, Gothenburg, Sept. 2017, paper W.3.A.1.
- [4] A. E. Willner and Syang-Myau Hwan, "Transmission of many WDM channels through a cascade of EDFA's in long-distance links and ring networks," in *Journal of Lightwave Technology*, vol. 13, no. 5, pp. 802-816, May 1995.
- [5] Y. Pointurier *et al.*, "Design of Low-margin Optical Networks," *Proc. OFC, Tu3F.5* (2016).
- [6] Rasmussen, Carl Edward, and Christopher KI Williams. "Gaussian processes in machine learning." *Lecture notes in computer science* 3176 (2004): 63-71.